Young Lives Methods Guide

Introducing Computer-Assisted Personal Interviewing

July 2011

Data for the first two rounds of the Young Lives survey were collected using paper-based questionnaires. Once completed, the questionnaires were manually checked, entered and then re-entered into databases, cross-checked to highlight inconsistencies, cleaned and then transferred into data analysis software. Since the start of the study in 2002, various Computer-Assisted Personal Interviewing (CAPI) options have become available for replacing much of this process with electronic data collection, which allows fieldworkers to go to research sites with questionnaires already loaded onto a laptop, notebook or personal digital assistant (PDA) onto which they can directly enter responses in a format that can be immediately uploaded into a database. This greatly increases the speed with which survey data can be accessed in a useable format, and means that some data inconsistencies can be tackled while fieldworkers are still with respondents in the research site.

In 2008, after Round 2 was completed, the many potential advantages of CAPI led Young Lives to explore the possibilities of using it in subsequent rounds. A fundamental decision made at this stage was that the complicated nature of the survey required bespoke software which would allow Young Lives staff to develop their own programmes for different questionnaire components. Young Lives hired a consultant to review the available software and hardware, and the findings of this review were presented at a meeting of the international team. This resulted in a decision to pilot PDAs and bespoke software in all four countries alongside the pilot of the Round 3 paper questionnaire. The tight scheduling of this process, which meant that questionnaires and software were being developed and tested simultaneously, led to some confusion and difficulties, but overall it was found that the technology was viable and the research teams willing to use it.

After this pilot stage, the Peruvian and Vietnamese study teams were keen to use electronic data collection for a proportion of the full Round 3 survey. PDAs were used to collect over 70% of the Round 3 data in Vietnam and 50% in Peru, and were also used for some questionnaires of the Schools Component in Ethiopia. Young Lives plans to use electronic data collection in all four countries in Rounds 4 and 5, and staff have embarked on a process of selecting the software and hardware needed to implement this.

Alongside the benefits that it brings, CAPI also presents challenges. CAPI technologies were developed for interviewing respondents in shopping malls and at trade shows in developed countries, and their transfer to often remote study sites with poor infrastructure demands careful consideration. In addition to this, the size and complexity of the Young Lives survey, with both country-specific components and multiple versions in different languages, presents particular challenges for software and programming.

Key considerations for implementing CAPI

Several different types of hardware – from hand-held devices such as mobile phones and PDAs to laptops and notebooks – can be used for electronic data collection. Diverse factors had to be taken into account when considering the best kind of hardware to carry out the survey:

- cost
- size. One researcher reflected that considering "the reality of sitting on a stone in the middle of a farmyard, balancing everything on your knee" was centrally important to evaluating the advantages and disadvantages of different kinds of hardware
- battery use and ease of battery charging in areas lacking reliable mains electricity
- impact on interview dynamics. Some fieldworkers voiced concerns about the respondent perceptions of different types of hardware in remote areas where technology can be viewed with extreme suspicion and fear by local people. Others also suggested that laptop screens would form a barrier between fieldworkers and respondents that might influence responses
- security and threat of theft
- software compatibility. Some of the software that Young Lives considered can only be used with a single type of hardware.

Having decided that the Young Lives survey would need programmable software, several other considerations needed to be taken into account when selecting software for electronic data collection:

- Internet connectivity. Some electronic data collection software is designed for use with the internet, and cannot function without a reliable, fast connection, which is not available in all study sites. Internet-based software does offer several advantages for the survey, including joint online working on programmes across countries and the opportunity to administer the questionnaire over the internet to some children who have moved and have e-mail addresses. Internet connectivity also facilitates the development of a version control system that tracks changes in files across the whole survey.
- Built-in checks. Different types of software have various levels of built-in checks which alert fieldworkers when there are possible inconsistencies in survey responses. Country teams had diverse views about the level of built-in checks they would find useful in the field.

- Opportunities for notes and comments. Some researchers emphasised the utility of being able to make written comments, particularly to qualify data which does not fall easily into existing codes or outside standard consistency checks.
- Language. A central feature of Young Lives is that it applies questionnaires in local languages. Different software present a range of approaches towards translation and surveys in different languages, ranging from translation functions to building separate versions of the programme for each country.
- Version control. As each round of the survey is designed, piloted and revised, multiple versions of the questionnaire are in circulation. Most software includes a version control system to provide a central coordinating function, and some will automatically update a change made in one country programme in other country programmes.

Planning electronic data collection for the Round 4 survey, scheduled for 2013, has involved balancing these diverse considerations. Options for programmable software having increased since the pilot stage, three companies were selected and invited to present their software to the Young Lives team. Each had both benefits and drawbacks, so making a final choice was a question of evaluating trade-offs. EDI, the software eventually selected, offered the simplest user interface and the most straightforward programming language, as well as dedicated technical support from a company experienced in using the software in developing countries. Disadvantages included no facility for internet-based version control, and the need to build five separate versions of the survey – a master programme with core questions, and four programmes for country-specific components in Amharic, Telugu, Spanish and Vietnamese. These drawbacks mean that special attention will have to be paid to developing and implementing management protocols about programming changes and consistency between versions.

EDI cannot be used with PDAs, and final decisions about hardware selection will be taken on a country-by-country and sometimes site-by-site basis. While data will be collected electronically in the vast majority of sites, security considerations mean that data could still be recorded on paper in a small minority of locations.

Challenges in implementing electronic data collection

Implementing CAPI for Round 4 will begin with training in the UK for all programmers, assistants and data managers. As questionnaire sections are finalised, they will be built into programmes, and the programming and questionnaire will be piloted together. Challenges remain in rolling out CAPI across the whole study, many of which intersect with broader issues of team management and training:

- Participation in programming. Having selected software that allows Young Lives data management staff to build programmes for data collection, there are trade-offs involved between strengthening ownership and capacity by involving multiple staff in programming, and minimising errors and confusion by involving fewer staff.
- Balancing checking and validation. As already noted, some teams want more automatic checks built into the data collection software which quickly alert fieldworkers to anomalies in the responses they are recording. The challenge here will be to find a balance between challenging the answers of respondents when they fall outside expected norms and addressing the tendency for some interviewers to want to provide clean-looking data that meets expectations.
- Potential bias. The possibility that using electronic data would have an influence on the quality of data in comparison to paper-based questionnaires. The Peru team tested a sample of the data collected during Round 3 for such bias, and found that electronically collected data had a high level of agreement with data collected using paper-based questionnaires (Escobal and Benites, forthcoming).
- Meeting country needs. Ensuring that programmes are built which meet the needs of each country team, especially as regards local languages, has to be balanced with centralised coordination of the whole survey across all four countries.
- Changing existing practices. During the first two rounds of the survey, when the questionnaires were still on paper, teams had become accustomed to being able to make changes and revisions until just before the start of fieldwork. This is much harder to achieve with CAPI, and changes in working practices will be needed to take this into account.
- Embracing new technologies. Many fieldworkers have been with Young Lives since the first round of the survey. They have diverse and valuable skills which range from fluency in different languages to the ability to motivate and win the respect of the children's families, to the perseverance and willingness to undertake difficult and sometimes dangerous journeys. These qualities do not necessarily go with CAPI skills, and this presents particular challenges for designing and delivering training.

References

Escobal J. and S. Benites (forthcoming, 2012) 'PDAs in Socio-economic Surveys: Instrument Bias, Surveyor Bias, Or Both? ', *International Journal of Social Research Methodology*